Generalized entropy models

Fosgerau, Mogens and de Palma, André

Technical University of Denmark, Denmark, Ecole Normale Supérieure de Cachan and Centre d'économie de la Sorbonne, and KU Leuven

27 February 2015

Generalized entropy models

Mogens Fosgerau[*]   André de Palma[†]

March 23, 2016

Abstract

We formulate a family of direct utility functions for the consumption of a differentiated good. This family is based on a generalization of the Shannon entropy. It includes dual representations of all additive random utility discrete choice models, as well as models in which goods are complements. Demand models for market shares can be estimated by plain regression, enabling the use of instrumental variables. Models for microdata can be estimated by maximum likelihood.

Keywords: market shares; product differentiation; discrete choice; duality; generalized entropy

JEL: D01, C25, L1

---

[*]Technical University of Denmark, Denmark, mf@transport.dtu.dk

[†]CES, ENS Cachan, CNRS, Université Paris-Saclay, Cachan, France, and CECO, Ecole Polytechnique, andre.depalma@ens-cachan.fr.

1 Introduction

We construct a family of direct utility functions that describe consumer demand for one unit of a differentiated good. A consumer with income y and consumption $q = (q_1, ..., q_J)$ of the differentiated good has utility $u(q, y) = \tau y + q \cdot v + \Omega(q)$, where $\tau > 0$ and $v = \tau(a - p)$ is quality minus price in utility units. The function Ω belongs to a family of generalized entropies, defined through a number of conditions as a generalization of the Shannon (1948) entropy; it is a concave function that expresses taste for variety while leading to a tractable link between consumption and utility. We find a new general structure for generalized entropy, which enables us to provide rules for constructing generalized entropies and a range of specific examples showing that generalized entropy may be used to generate rich patterns of substitution and complementarity.

We share the idea of using convex analysis and duality in a discrete choice context with other recent contributions. Salanié and Galichon (2015) consider matching models with transferable utility and arrive at a generalization of entropy that belongs to our family of generalized entropies. Chiong et al. (2015) apply similar ideas to dynamic discrete choice models. Melo (2012) uses duality to show existence of a representative agent for a dynamic discrete choice model on a network. The essential contribution of this paper is the finding that generalized entropy has a certain structure that allows us to access a new and rich universe of tractable demand models that has not been explored before.

Models specified in terms of generalized entropy may be estimated using simple regression with instruments that are available within the model. In this respect, our paper is closely related to Berry (1994) and Berry and Haile (2014) who invert the market shares of an additive random utility model (ARUM) to find corresponding utility levels. Given that this transformation is known, Berry (1994) shows how model parameters may be estimated using standard instrumental variable regression techniques with inverted markets shares as dependent variables. Inversion of market shares may be carried out with an explicit formula for the case of the multinomial and the nested logit models. However, these models imply substitution patterns that may be implausible in many applications (Berry et al., 1995). More flexible substitutions patterns may be allowed using random pa-

rameter models, but then numerical methods are necessary to carry out the Berry inversion, which leads to numerical and computational issues in combination with the random parameters (Knittel and Metaxoglou, 2014).

In this paper we formulate models, not in the space of indirect utilities of discrete choice models, but in the dual space of consumption shares. This makes the inverted market shares directly available and numerical methods are unnecessary for computing them. Consistency with maximization of a well-behaved utility function is automatically ensured. We provide a range of examples leading to substitution and complementarity patterns that go well beyond the nested logit example. These may potentially be used as alternatives to the random coefficient logit in what has become known as Berry, Levinsohn, Pakes or just BLP models (Berry et al., 1995).

Our generalized entropy models can also be applied to microdata of discrete choices, allowing individual level information to be taken into account. In this case, numerical methods are required to compute the likelihood. The likelihood can be computed via a fixed point iteration that we show is guaranteed to converge in a range of circumstances. Then models can be estimated using maximum likelihood. Random parameters are not required to allow for more complex substitution patterns than plain or nested logit.

The family of models based on generalized entropy is large: we show that it comprises models corresponding to any ARUM. For the multinomial logit model, the corresponding generalized entropy is the Shannon entropy (Anderson et al., 1988). The generalized entropy family is in fact larger than the family of ARUM, we show that generalized entropies exist that lead to demands that are not consistent with any ARUM. Importantly, generalized entropy models exist where goods may be complements rather than substitutes, whereas goods are always substitutes in ARUM.

McFadden (1978) developed a family of discrete choice models based on the form of the expected maximum utility function when random utilities follow a multivariate extreme value (MEV) distribution. This family includes the multinomial and the nested logit models as the simplest special cases. McFadden (1978) applied a nesting device to utilities to create a range of instances of MEV models; in the present paper we create instances of generalized entropy models by

3

applying a nesting device to market shares.

Fudenberg et al. (2014) analyzes utility of the same form as used in this paper, but where the entropy term $\Omega(q)$ is separable as a sum of terms $f_j(q_j)$. It is crucial for the results in this paper not to require such separability. Mattsson and Weibull (2002) have a similar setup, but where $\Omega(q)$ is interpreted as an implementation cost and where axioms are imposed that essentially reduce $\Omega(q)$ to the Shannon entropy such that demand arises that is consistent with the logit model. This paper uses generalized entropy to describe substitution and complementarity patterns that go well beyond this.

The budget set for the consumer in this paper incorporates a quantity constraint and is hence not linear in income and prices. This fits into the framework of Fosgerau and McFadden (2012) who develop a micro-economic theory of consumer demand under general budgets and where utility is perturbed by a linear term such as $q \cdot v$.

Section 2 introduces generalized entropy and uses it to define and solve a class of direct utility models for market shares. A range of results and accompanying examples are presented that allows members of this class to be constructed. Section 3 shows how utility parameters in generalized entropy models may be recovered from market level data using standard regression techniques. Section 4 relates generalized entropy to discrete choice models and shows that all ARUM are represented by generalized entropy via duality. Section 4.1 presents a fixed point iteration that converges to the probability vector associated with utility levels v in a discrete choice setting and applies this in an example of maximum likelihood estimation using microdata of discrete choices. Section 5 concludes. Proofs are in the appendix.

2   Direct utility models for market shares

2.1   Notational conventions

Vectors are denoted simply as $q = (q_1, ..., q_J)$. A univariate function applied to a vector is understood as coordinate-wise application of the function, e.g., $e^q = (e^{q_1}, ..., e^{q_J})$. Consequently, if a is a real number then $a + q = (a + q_1, ..., a + q_J)$.

4

The multivariate function $S : \mathbb{R}^J \to \mathbb{R}^J$ is composed of univariate functions with superscripts (j): $S(q) = \left(S^{(1)}(q), ..., S^{(J)}(q)\right)$. Subscripts denote partial derivatives, e.g. $G_j(v) = \frac{\partial G(v)}{\partial v_j}$. The gradient with respect to a vector v is ∇_v; e.g., for $v = (v_1, ..., v_J)$, $\nabla_v G(v) = \left(\frac{\partial G(v)}{\partial v_1}, ..., \frac{\partial G(v)}{\partial v_J}\right)$. The Jacobian is denoted J with, for example,

$$J_{\ln S}(q) = \begin{pmatrix} \frac{\partial \ln S^{(1)}}{\partial q_1} & \cdots & \frac{\partial \ln S^{(1)}}{\partial q_J} \\ \cdots & \cdots & \cdots \\ \frac{\partial \ln S^{(J)}}{\partial q_1} & \cdots & \frac{\partial \ln S^{(J)}}{\partial q_J} \end{pmatrix}.$$

A dot indicates an inner product or products of vectors and matrixes. The unit simplex in $\mathbb{R}^J$ is Δ. A subset $g \subseteq \{1, ..., J\}$ is called a nest and we use the notation $q_g = \sum_{j \in g} q_j$ as shorthand for the sum of q over a nest g.

2.2 Consumer demand

Consider a consumer with income y facing a price vector p for J varieties of a differentiated good and a numeraire good with price 1. The consumer maximizes utility $\tau z + \tau q \cdot a + \Omega(q)$, where $\tau > 0$, q is the vector of quantities of the differentiated good, and z is the quantity of the numeraire good. The consumer has a budget constraint $y \geq z + q \cdot p$. Importantly, the consumer also has a quantity constraint $\sum_j q_j = 1$, which normalizes demand for the differentiated good. Income is sufficiently large, $y > \max_j \{p_j\}$, that consumption of the numeraire good is always positive. The budget constraint is always binding and substituting it into utility leads to

$$u(q, y) = \tau y + q \cdot v + \Omega(q), \tag{1}$$

where $v = \tau(a - p)$.

We begin by giving an abstract formulation of Ω; specific examples will be provided afterwards. *Generalized entropy* is a function $\Omega : [0, \infty)^J \to \mathbb{R} \cup \{-\infty\}$ given by

$$\Omega(q) = \begin{cases} -q \cdot \ln S(q), & q \in \Delta \\ -\infty, & q \notin \Delta \end{cases}, \tag{2}$$

where the function $S : [0, \infty)^J \to [0, \infty)^J$ is a *flexible generator*, defined next.

Note that the domain of generalized entropy embodies the constraint that demands q_j sum to 1.[1]

A function S is a *flexible generator* if it satisfies the following four conditions.

Condition 1 *S is continuous, and homogenous of degree 1.*

Condition 2 *Ω is concave.*

Condition 3 *S is differentiable at any $q \in \text{relint}(\Delta)$ with*

$$\sum_{j=1}^{J} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k} = \kappa, k \in \{1, ..., J\},$$

where $\kappa > 0$.

Condition 4 *S is globally invertible.*

In order to build intuition, let us consider what happens if the components $S^{(j)}$ of a flexible generator are identical and, as in Fudenberg et al. (2014), each $S^{(j)}$ depends only on q_j. Then Condition 3, which may be expressed as $q \cdot J_{\ln S}(q) = \kappa(1, ..., 1)$, reduces to $\frac{\partial \ln S^{(j)}(q_j)}{\partial q_j} = \kappa/q_j$, which implies that $S^{(j)}(q_j) = cq_j$, for some $c > 0$. The function $S(q) = cq$ satisfies Conditions 1-4 and the corresponding generalized entropy $\Omega(q) = -q \cdot \ln q - \ln c$ is just the Shannon entropy up to a constant. Maximizing utility (1) with this entropy under the quantity constraint $\sum_j q_j = 1$ leads to logit demand (Anderson et al., 1988)

$$q(v) = \left(\frac{e^{v_1}}{\sum_{j=1}^{J} e^{v_j}}, ..., \frac{e^{v_J}}{\sum_{j=1}^{J} e^{v_j}} \right).$$

In general, each $S^{(j)}$ depends on the whole vector q, which complicates the derivation of an expression for the demand. Here Condition 3 plays a key role, ensuring that $-\partial \Omega(q)/\partial q_k = \ln S^{(k)}(q) + \kappa$. This leads to a tractable and familiar form for demand as shown in the next theorem.

---

[1] We will show (Theorem 4) that the convex conjugate of the ARUM surplus function has this form.

Theorem 1 *Let Ω be a generalized entropy as given in ([2](#)). Maximization of utility $u(q, y) = \tau y + q \cdot v + \Omega(q)$ leads to a demand system with interior solution*

$$q(v) = \left(\frac{H^{(1)}(e^v)}{\sum_{j=1}^{J} H^{(j)}(e^v)}, ..., \frac{H^{(J)}(e^v)}{\sum_{j=1}^{J} H^{(j)}(e^v)} \right), \tag{3}$$

where $H = S^{-1}$.

Demand q corresponds to v in the expression ([3](#)) if and only if v and q are related through the flexible generator S by $v = \ln S(q) + c$ for some $c \in \mathbb{R}$.

As we have seen, the form ([3](#)) of demand generalizes the logit demand. We shall establish in Section [4](#) that for any ARUM there exists a generalized entropy that leads to the same demand. We shall also show in Theorem [2](#) that generalized entropies exist that are not consistent with ARUM demand.

The second part of Theorem [1](#) establishes that utility can be computed up to a constant directly from demand, given a flexible generator S. This result is used in Section [3](#), which discusses estimation of these models via regression.

Throughout the paper, we denote the inverse of a flexible generator S by $H \equiv S^{-1}$. The formulation of generalized entropy does not rule out corner solutions in general. Whether zero demands can arise depends on the specific formulation of generalized entropy.

We end this section by a proposition, proved in Fosgerau and McFadden (2012)[2], showing that each demand q_j is weakly increasing as a function of the corresponding v_j. More generally, it establishes a cyclical monotonicity condition (Rockafellar, 1970, chap. 24) which guarantees that demand is contained in the subdifferential of a convex function.

Proposition 1 (Cyclical monotonicity) *If $\left\{v^k\right\}_1^{K+1}, K \geq 1$ is a finite sequence of vectors with $v^{K+1} = v^1$, and demand q is as described in Theorem [1](#), then*

$$\sum_{k=1}^{K} \left(v^{k+1} - v^k \right) \cdot q\left(v^k \right) \leq 0. \tag{4}$$

Each demand function $q_j(v)$ is weakly increasing in $v_j, j = 1, ..., J$.

---

[2]We have not been able to find an earlier statement of this result.

We proceed to construct instances of generalized entropy for applications.

2.3 Construction of generalized entropies

We have already identified one flexible generator, namely the identity $S(q) = q$. The following subsections provide ways to generate many more flexible generators. An obstacle that we will face is to establish invertibility of candidate flexible generators. To overcome this, we have the following lemma, adapted from a global inversion theorem for homogeneous nonlinear maps.

Lemma 1 (Ruzhansky and Sugimoto 2014) *Let $J \geq 3$ and let $S\colon (0, \infty)^J \to (0, \infty)^J$ be continuously differentiable, linearly homogenous with a Jacobian determinant that never vanishes and with $\inf_{q \in \Delta} \|S(q)\| > 0$. Then S is invertible.*

In the examples below we will see ways to construct functions that satisfy Conditions 1-3. In order for these functions to be flexible generators, it then remains to ensure that they are invertible. Building on Lemma 1, the next lemma establishes conditions under which the weighted geometric average of such functions, where just one of them must itself be a flexible generator, leads to a new flexible generator.

Lemma 2 (Averaging) *Let $T_1, ..., T_K : (0, \infty)^J \to (0, \infty)^J$ satisfy Conditions 1-3, where the Jacobian of each $\ln T_k$ is symmetric and positive semidefinite and positive definite for at least one k. If $T_k^{(j)}(q) \geq q_j$ for each k and j and $\alpha_1, ..., \alpha_K$ are positive numbers that sum to 1, then $S\colon (0, \infty)^J \to (0, \infty)^J$ given by*

$$S = \prod_{k=1}^{K} T_k^{\alpha_k}$$

is a flexible generator.

As a consequence, a mapping created by averaging the identity $T_1(q) = q$ with some T_2 that satisfies the conditions of the lemma except positive definiteness is always invertible and hence it is a flexible generator.

Proposition 2 presents a general construction of flexible generators through a nesting operation. A nest g is a set of goods for which a term $q_g^{\mu_g}$ enters the
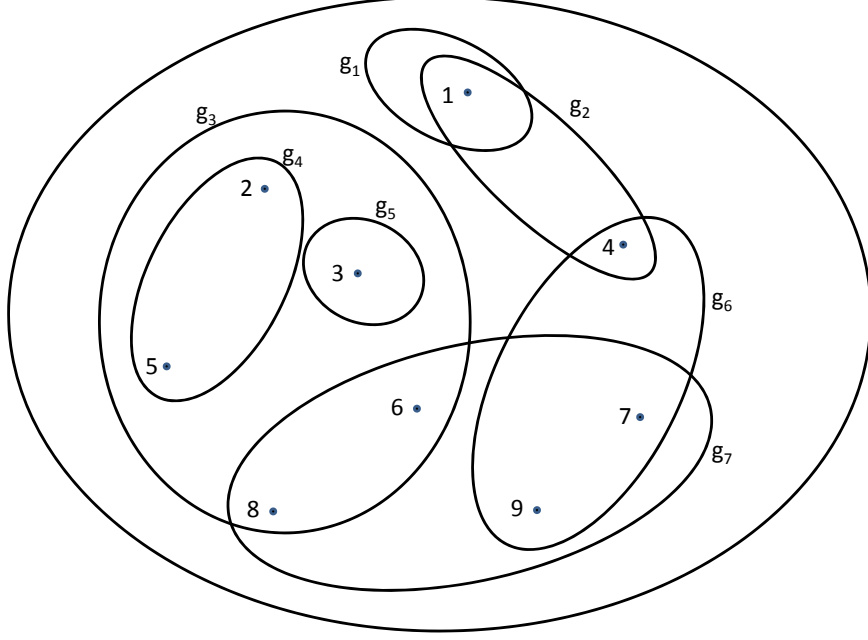
Figure 1: Nesting example with 9 goods and 7 nests.

entropy component of utility, where $\mu_g \in \,]0,1]$ is a nesting parameter. The closer μ_g is to 1, the more the goods in nest g act in the utility as one single good and they become closer to being perfect substitutes. The division of alternatives into nests is illustrated in Figure 1. As the figure shows, one alternative may belong in several nests, and nests may or may not be subsets of other nests. Proposition 2 requires that the nesting parameters sum to 1, summed across the nests that contain any given of the J goods.[3]

Proposition 2 (General nesting) *Let $\mathcal{G} \subseteq 2^{\{1,...,J\}}$ be a finite set of nests with associated nesting parameters μ_g, where $\sum_{\{g \in \mathcal{G} | j \in g\}} \mu_g = 1$ for all j and $\mu_g > 0$ for all $g \in \mathcal{G}$. Let $S = \left(S^{(1)}, ..., S^{(J)}\right)$ be given by*

$$S^{(j)}(q) = \prod_{\{g \in \mathcal{G} | j \in g\}} q_g^{\mu_g}. \tag{5}$$

---

[3]In the example this may achieved by letting $\mu_1 = \mu_3 = \mu_6 = \mu > 0$ and $\mu_2 = \mu_4 = \mu_5 = \mu_7 = 1 - \mu > 0$.

9

Then S satisfies Conditions 1-3, the Jacobian of $\ln S$ is symmetric and positive semidefinite, and for each j, $S^{(j)}(q) \geq q_j$. If the Jacobian of $\ln S$ is positive definite, then S has an inverse and S is a flexible generator.

The following examples illustrate the application of Proposition 2 to construct a flexible generator.

Example 1 *Consider $J \geq 3$ with all possible nests with 1 or 2 alternatives as elements, e.g. for $J = 3$:*

$$\mathcal{G} = \{\{1\}, \{2\}, \{3\}, \{1, 2\}, \{1, 3\}, \{2, 3\}\}.$$

Each alternative belongs to J nests and we let $\mu_g = 1/J$. Define in accordance with (5) the function S by

$$S^{(j)}(q) = q_j^{\frac{1}{J}} \prod_{i \neq j} (q_i + q_j)^{\frac{1}{J}}.$$

By Proposition 2 this is a flexible generator. The demand solves $S(q) = e^{v-c}$ for some $c \in \mathbb{R}$.

The next example shows that Proposition 2 leads to the nested logit model as a special case.

Example 2 *Partition the set of alternatives $\{1, ..., J\}$ into nests $g \in \mathcal{G}$ and denote by g_j the nest that contains alternative j. Let*

$$S^{(j)}(q) = q_j^{\mu_{g_j}} q_{g_j}^{1-\mu_{g_j}}, j \in g_j, \tag{6}$$

where $\mu_{g_j} \in]0, 1]$ are parameters. Then S is a flexible generator by Proposition 2. It is straightforward to verify that the equation $S(\tilde{q}) = e^v$ has solution

$$\tilde{q}_j = e^{\frac{v_j}{\mu_{g_j}}} \left(\sum_{i \in g_j} e^{\frac{v_i}{\mu_{g_j}}} \right)^{\mu_{g_j}-1}.$$

10

Normalizing the sum of demands to 1 *leads to*

$$q_j = \frac{\tilde{q}_j}{\sum_{g \in \mathcal{G}} \tilde{q}_g} = \frac{e^{\frac{v_j}{\mu_{g_j}}}}{\sum_{i \in g_j} e^{\frac{v_i}{\mu_{g_j}}}} \frac{e^{\mu_{g_j} \ln\left(\sum_{i \in g_j} e^{\frac{v_i}{\mu_{g_j}}}\right)}}{\sum_{g \in \mathcal{G}} e^{\mu_g \ln\left(\sum_{i \in g} e^{\frac{v_i}{\mu_g}}\right)}},$$

which is the nested logit model (McFadden, 1978).[4]

We shall now use the general nesting result of Proposition 2 to create a cross-nested model, which generalizes the nested logit model. Say that a set of products can be naturally grouped according to two criteria, where one grouping is not a subdivision of the other. For example, automobiles may be grouped according to brand or according to body type. We shall create a structure that is similar to the nested logit model, but which, unlike the nested logit model, allows for non-nested groupings.[5] In this example, we also include an outside good, with index zero.

Example 3 *Let* $\mu_0, \mu_1, \mu_2 > 0$, $\mu_0 + \mu_1 + \mu_2 = 1$. *Let* $\sigma_c(j)$ *be the set of products that are grouped together with product* j *on criteria* $c = 1, 2$. *Denote as before* $q_{\sigma_c(j)} = \sum_{i \in \sigma_c(j)} q_i$ *and define* S *by*

$$S^{(j)}(q) = \begin{cases} q_0, & j = 0 \\ q_j^{\mu_0} q_{\sigma_1(j)}^{\mu_1} q_{\sigma_2(j)}^{\mu_2}, & j > 0. \end{cases} \tag{7}$$

Then it follows directly from Proposition 2 that S *is a flexible generator. The cross-nesting model is applied in Section 3.1.*

The next proposition provides a case that goes beyond averaging of simple nesting flexible generators and where the inversion of market shares can be carried out to yield a closed form expression for demand.

---

[4]Berry (1994) noticed the explicit inversion of the nested logit demand and used inversion of market shares to estimate utility parameters using standard regression techniques. Verboven (1996) used the same inversion when deriving nested logit demand for a representative consumer.

[5]With only the nested logit model available, researchers have been forced to choose a hierarchy of criteria, for example first grouping cars by make and then by body type within each make. With cross-nesting, it is not necessary to fix such hierarchy.

Proposition 3 (Invertible nesting) *Let S be given by (5), where the number of nests is equal to the number of alternatives. Let $W = diag\left(\mu_{g_1}, .., \mu_{g_J}\right)$ be a diagonal matrix of positive nesting parameters and let $M_{J \times J} = \left\{1_{\{j \in g\}}\right\}$ be an incidence matrix, where rows correspond to alternatives and columns correspond to nests. Suppose that M is invertible. Then S has an inverse and S is a flexible generator. Moreover, unnormalized demand satisfies*

$$v = \ln S\left(\tilde{q}\right) \Leftrightarrow \tilde{q} = \left(M^\top\right)^{-1} \exp\left(W^{-1} M^{-1} v\right).$$

The next example illustrates the application of Proposition 3.

Example 4 *Consider $J \geq 3$ and define nests from the symmetric incidence matrix M with entries $M_{ij} = 1_{\{i \neq j\}}$. Then each alternative is in $J - 1$ nests and we may associate weights $\mu_g = 1/(J - 1)$ with each nest. The inverse of the incidence matrix has entries $\left(M^{-1}\right)_{ij} = \frac{1}{J-1} - 1_{\{i=j\}}$. Solving $\ln S\left(\tilde{q}\right) = v$ leads to $\tilde{q} = M^{-1} \exp\left[(J - 1) M^{-1} v\right]$,or equivalently*

$$
\begin{aligned}
\tilde{q}_i &= \sum_{j=1}^{J} \left(\frac{1}{J-1} - 1_{\{i=j\}}\right) \exp\left(\sum_{k=1}^{J} \left(1 - (J-1) 1_{\{k=j\}}\right) v_k\right) \\
&= \sum_{j=1}^{J} \left(\frac{1}{J-1} - 1_{\{i=j\}}\right) \exp\left(\sum_{k=1}^{J} v_k\right) e^{-(J-1)v_j} \\
&= \exp\left(\sum_{k=1}^{J} v_k\right) \left(\frac{1}{J-1} \sum_{j=1}^{J} e^{-(J-1)v_j} - e^{-(J-1)v_i}\right).
\end{aligned}
$$

Normalized demand is then

$$q_i = \frac{\sum_{j=1}^{J} e^{-(J-1)v_j} - (J-1) e^{-(J-1)v_i}}{\sum_{j=1}^{J} e^{-(J-1)v_j}}.$$

The model in the previous example looks similar to the multinomial logit but is different in important ways. First, it does not have the independence from irrelevant alternatives property. Second, zero demands may arise.[6] The above expression for demand leads to non-negative demands only for values of v within

---

[6]Zero demands may also arise in an ARUM where the error terms have bounded support.

some set. A way to ensure that demands are strictly positive is to average with a flexible generator such as the simple identity, since then $\ln q_j$ must all be finite.

Third, the demand from the invertible nesting model in the example is not consistent with any ARUM. ARUM demand has the restrictive feature that the mixed partial derivatives of q_j alternate in sign (McFadden, 1981). This feature is not exhibited by the demand generated in this example, since $\frac{\partial q_1}{\partial v_2} < 0, \frac{\partial^2 q_1}{\partial v_2 \partial v_3} < 0.$[7] Thus, we have established the following theorem.

Theorem 2 *Some generalized entropies lead to demand systems that cannot be rationalized by any ARUM.*

In Section 4 we establish that all ARUM have a generalized entropy as counterpart that leads to the same demand. Thus the class of generalized entropies is strictly larger than the class of ARUM models.

The signs of the mixed partial derivatives of a quantity with respect to the prices of other goods vary in the same way also for CES demand under the standard linear budget constraint when CES utility is $u(x) = \sum_{j=1}^{J} \alpha_j x_j^{\gamma}, \alpha_j > 0, \gamma \in (0, 1)$. It is thus possible for a well-behaved utility function that the signs of the mixed partial derivatives of q_j are not consistent with those predicated by ARUM.

Consider now a pair Ω, S of generalized entropy and flexible generator. If A is a $J \times J$ permutation matrix, then $q \to \Omega(Aq)$ is also a generalized entropy, since application of a permutation matrix to q just amounts to a reordering of the dimensions of q. The convex hull of the set of $J \times J$ permutation matrixes is the set of $J \times J$ doubly stochastic matrixes, i.e. matrixes with non-negative elements that sum to 1 across rows and columns (Birkhoff, 1946; Mirsky, 1958) The following proposition shows more generally how a flexible generator can be transformed into a new flexible generator by a location shift and a matrix with non-negative entries that sum to 1 across columns.

---

[7]Note that $\frac{\partial q_1}{\partial v_2} \sim -(J-1)^2 e^{-(J-1)(v_1+v_2)} < 0$ and $\frac{\partial^2 q_1}{\partial v_2 \partial v_3} \sim -2(J-1)^3 e^{-(J-1)(v_1+v_2+v_3)} < 0.$

Proposition 4 (Transformation) *Let T be a flexible generator, $m \in \mathbb{R}^J$, and let $A = \{a_{ij}\} \in \mathbb{R}^J \times \mathbb{R}^J$ be invertible with $a_{ij} \geq 0$ and $\sum_i a_{ij} = 1$. Then*

$$S : q \rightarrow \exp\left(A^\top \left[\ln\left(T\left(Aq\right)\right)\right] + m\right) \tag{8}$$

is a flexible generator.

We shall illustrate Proposition 4 with a flexible generator that leads to demand where goods may be complements in the sense that the demand for one good increases as the utility component v_j of another good increases.

Example 5 *Let $J = 3$ and define*

$$A = \begin{pmatrix} .4 & .6 & 0 \\ .6 & .4 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Compute demand according to Proposition 4 with $m = 0$ to find that

$$\tilde{q} = A^{-1}\left(\exp\left[\left(A^\top\right)^{-1} v\right]\right) = \begin{pmatrix} 3e^{3v_1 - 2v_2} - 2e^{3v_2 - 2v_1} \\ 3e^{3v_2 - 2v_1} - 2e^{3v_1 - 2v_2} \\ e^{v_3} \end{pmatrix},$$

which leads to $q_3 = \frac{e^{v_3}}{e^{3v_2 - 2v_1} + e^{3v_1 - 2v_2} + e^{v_3}}$. Then $\frac{\partial q_3}{\partial v_1} > 0$ iff $v_2 - v_1 > \frac{1}{5} \ln \frac{3}{2}$.

Goods are always substitutes in an ARUM. Complementarity is, however, important for describing situations where some goods tend to be bought together, for example taco chips and salsa. The example above establishes that generalized entropy models are able to allow goods to be complements. We state this insight as a theorem and note that this is also another example of a generalized entropy model that is not consistent with any ARUM.

Theorem 3 *Some generalized entropies allow goods to be complements.*

The last proposition in this section presents a nesting device that can be used to combine flexible generators into new flexible generators.

Proposition 5 (Nesting) *Let T_1, T_2 be flexible generators with $T_1 : \mathbb{R}^{J_1} \to \mathbb{R}^{J_1}$ and $T_2 : \mathbb{R}^{J_2} \to \mathbb{R}^{J_2}$. Then $S : \mathbb{R}^{J_1+J_2-1} \to \mathbb{R}^{J_1+J_2-1}$ defined for $q^1 \in \mathbb{R}^{J_1}$ and $q^2 \in \mathbb{R}^{J_2-1}$ by*

$$S^{(j)}\left(q^1, q^2\right) = \begin{cases} T_1^{(j)}\left(\frac{q^1}{1 \cdot q^1}\right) T_2^{(1)}\left(1 \cdot q^1, q^2\right), & j \leq J_1 \\ T_2^{(j-J_1)}\left(1 \cdot q^1, q^2\right), & J_1 < j \leq J_1 + J_2 - 1 \end{cases} \tag{9}$$

is a flexible generator with inverse given by $H\left(e^{v^1}, e^{v^2}\right) = \left(sT_1^{-1}\left(e^{v^1}\right), q^2\right)$, where s is given by $\left((1 \cdot q^1) s, q^2\right) = T_2^{-1}\left((1 \cdot q^1), e^{v^2}\right)$.

Propositions 2-5 allow a wide range of flexible generators to be constructed for applications. Through averaging and nesting operations it is possible to combine patterns of substitution and complementarity in a single model.

3 Estimation of generalized entropy models

In this section we consider the estimation of generalized entropy models from market share data. Later, in Section 4.1, we consider estimation based on micro data of discrete choices.

Flexible generators may be used to estimate market share models in a way similar to Berry (1994). Berry starts from the perspective of a discrete choice model and inverts market shares to determine utility levels (up to a constant) associated with a set of products in a number of markets. These utility levels form the basis for a regression where instrumental variable techniques may be used to deal with endogeneity, notably occurring if there are unobserved quality attributes that are correlated with prices. Here we shall exploit Theorem 1, which delivers utility levels (up to a constant) as a flexible generator applied to a vector of market shares. Models specified in terms of flexible generators thus circumvent the need to invert market shares numerically, while offering the opportunity to use functional forms that generalize the nested logit model.

Let us consider a market with J products and an outside good. The market share q_j of product j depends only on utility levels $v = (v_1, ..., v_J)$, where $v_j = z_j \cdot \beta + \xi_j$. The ξ_j is an unobserved demand characteristic of product j, which is

mean independent of z and independent across markets, z_j is a vector of variables and β is a vector of parameters to be estimated. The utility of the outside good is normalized as $v_0 = 0$. Assume further that demand given v is (3), where H is the inverse of a flexible generator S. Then, by Theorem 1, we have $\ln S(q) = v + c$, where $c \in \mathbb{R}$, or equivalently

$$\ln S^{(j)}(q) - \ln S^{(0)}(q) = z_j \cdot \beta + \xi_j. \tag{10}$$

Given a specific form for S, (10) may be estimated using linear regression techniques. Given suitable instruments, it is possible to allow for endogeneity of some of the variables in z_j. Here we shall focus on the estimation of the parameters in $\ln S^{(j)}$. We shall provide two examples: the first has a cross-nested structure, the second has an ordered structure.

3.1 A cross-nested model for market shares

We consider the cross-nesting example 3. Cross-nesting is appropriate if there are several dimensions along which products may be similar and closer substitutes for each other. We have mentioned the example of automobiles.

Insert (7) into (10), rearrange slightly and reparametrize using $\tilde{\beta} = \frac{\beta}{\mu_0}, \tilde{\mu}_1 = \frac{\mu_1}{\mu_0}, \tilde{\mu}_2 = \frac{\mu_2}{\mu_0}, \delta = \frac{1}{\mu_0}, \tilde{\xi}_j = \frac{1}{\mu_0}\xi_j$ to obtain the regression

$$\ln q_j = z_j \cdot \tilde{\beta} - \tilde{\mu}_1 \ln q_{\sigma_1(j)} - \tilde{\mu}_2 \ln q_{\sigma_2(j)} + \delta \ln q_0 + \tilde{\xi}_j. \tag{11}$$

This can be estimated treating $\ln q_{\sigma_1(j)}, \ln q_{\sigma_2(j)}$ and $\ln q_0$ as endogenous. Potential instruments include characteristics of products i that share nests with product j as well as the sum of characteristics over all products.

We have simulated data for this model using a cross-nested structure as shown in Figure 2. There are three by three alternatives and an outside option. There is one explanatory variable z_j, which is i.i.d. standard normal. Unobserved characteristics $\tilde{\xi}_j$ are i.i.d. standard normal multiplied by a factor 1/2. We set $(\beta, \mu_1, \mu_2) = (1, 0.1, 0.4)$, such that there is both a small and a larger nesting parameter. True regression parameters become these divided by $1 - \mu_1 - \mu_2$. The market shares $(q_0, q_1, ..., q_9)$ corresponding to each draw of $(z_1, ..., z_9)$ and
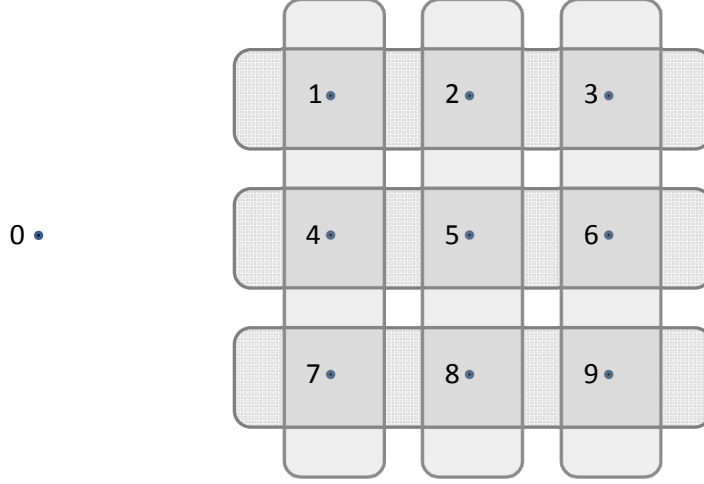
16

Figure 2: Cross-nested structure of model in the simulation example, with 3 by 3 products and an outside option 0.

$\left(\tilde{\xi}_1, ..., \tilde{\xi}_9\right)$ are determined by solving numerically the utility maximization problem in Theorem 1. We have generated 1000 datasets with 100 observations in each, where one observation consists of vectors $(q_0, q_1, ..., q_9)$ and $(z_1, ..., z_9)$.

For each dataset we estimate the regression (11) using instrumental variable (IV) regression with instruments $1, z_j, z_{\sigma_1(j)}, z_{\sigma_2(j)}, \sum_i z_i$ and squares of these. These instruments correlate with the endogenous variables and are independent of the noise term by construction of the data. F-statistics for the excluded instruments in the first-stage regression range mostly above 100 for $\ln q_{\sigma_1(j)}$ and $\ln q_{\sigma_2(j)}$. For $\ln q_0$, F-statistics are lower but still with average around 100 and minimum above 30.

Table 1 summarizes the simulation. The average of the IV estimates is close to the true values; the corresponding standard deviations may be considered small considering that each dataset only has 100 observations. The average OLS estimates are all more than two standard deviations from their true values, which indicates that the instruments play a significant role in the IV estimation.

17

Table 1: Parameter estimates in simulation with cross-nested model

	$\widetilde{\beta}$	$-\widetilde{\mu}_1$	$-\widetilde{\mu}_2$	δ
True parameters	2	-0.2	-0.8	2
Avg. IV estimates	2.00	-0.20	-0.79	1.99
Std.dev.	0.04	0.05	0.08	0.06
Avg. OLS estimates	1.76	0.10	-0.41	1.59
Std.dev.	0.04	0.04	0.05	0.05

3.2 An ordered model for market shares

The cross-nested model that we estimated in the previous section is among the simplest of the new models that we can create using flexible generators. Many more models can be created using Proposition 2. We shall now present an example where there is an ordering among products such that products that are nearer each other in the ordering are closer substitutes.

Products $1, ..., J$ are ordered in sequence. For simplicity, the ordering is circular such that there are no endpoints. There is an outside option 0 with markets share q_0. Define a flexible generator S by

$$S^{(j)}(q) = \begin{cases} q_0, & j = 0 \\ q_j^{\mu_0} I_1^{\mu_1}(j) I_2^{\mu_2}(j) I_3^{\mu_3}(j), & j > 0, \end{cases}$$

where $I_1(j) = q_{j-2} + q_{j-1} + q_j$, $I_2(j) = q_{j-1} + q_j + q_{j+1}$, $I_3(j) = q_j + q_{j+1} + q_{j+2}$ and parameters μ_i are positive and sum to 1. This is a flexible generator by Proposition 2. The structure is illustrated in Figure 3. There is a nest for any triple of neighboring products and each product is then in three nests. Then each product has its immediate neighbors as closest substitute and next neighbors as less close substitutes.

As before we simulated 1000 datasets from this model with 100 observations in each dataset. Variables z_j and $\tilde{\xi}_j$ are again respectively i.i.d. $N(0,1)$ and i.i.d. $0.5 \cdot N(0,1)$. We estimate the regression,
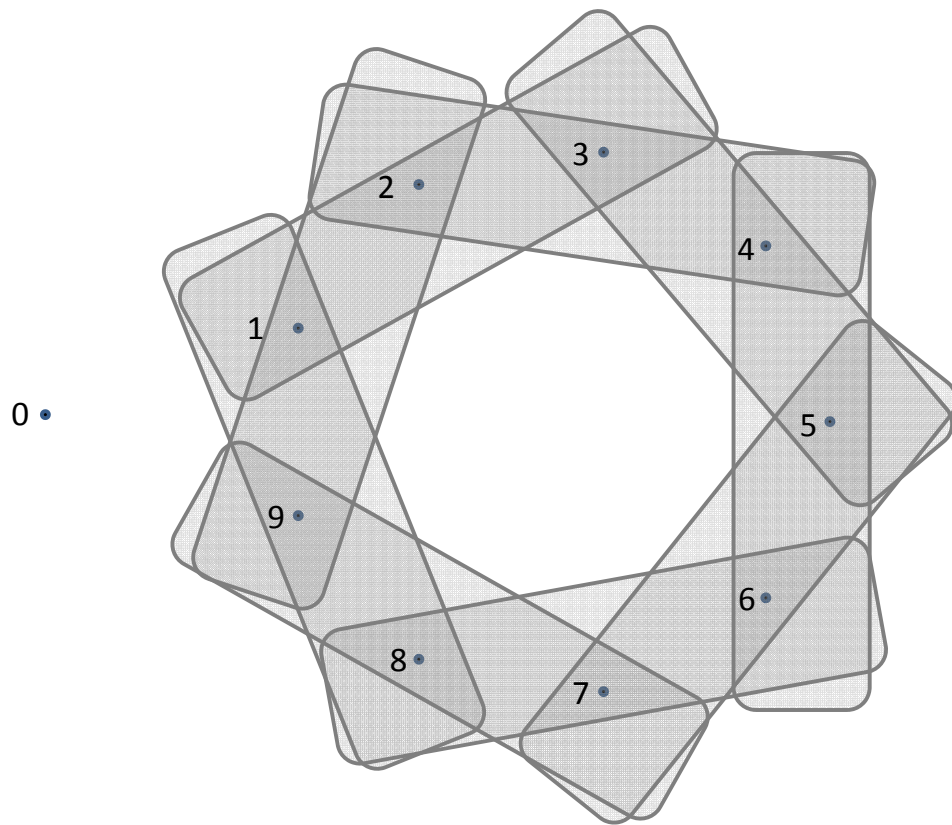
Figure 3: Ordered structure of model in simulation example products and an outside option

Table 2: Parameter estimates in simulation with ordered model

	$\widetilde{\beta}$	$-\widetilde{\mu}_1$	$-\widetilde{\mu}_2$	$-\widetilde{\mu}_3$	δ
True parameters	2.50	-0.50	-0.50	-0.50	2.50
Avg. IV estimates	2.49	-0.49	-0.49	-0.49	2.49
Std.dev.	0.06	0.08	0.08	0.08	0.08
Avg. OLS estimates	2.16	-0.10	-0.36	-0.10	1.91
Std.dev.	0.06	0.05	0.06	0.06	0.06

$$
\begin{aligned}
\ln q_j &= z_j \cdot \tilde{\beta} - \tilde{\mu}_1 \ln \left(\sum_{j-2 \leq i \leq j} q_i \right) - \tilde{\mu}_2 \ln \left(\sum_{j-1 \leq i \leq = j+1} q_i \right) \\
&\quad - \tilde{\mu}_3 \ln \left(\sum_{j \leq i \leq j+2} q_i \right) + \delta \ln q_0 + \tilde{\xi}_j,
\end{aligned}
$$

using the same transformation of parameters as before. Note that we allow for three different values of $\tilde{\mu}_i$, although they all have the same true value $\tilde{\mu}_i = \mu_1/\mu_0$. As instruments we use $1, z_j, \sum_{j-2 \leq i \leq j} z_i, \sum_{j-1 \leq k \leq = j+1} z_i, \sum_{j \leq k \leq j+2} z_i$ as well as squares of these variables. F-statistics for the excluded instruments in the first-stage regression are again very high.

Estimation results are summarized in Table 2. As before, the average of the IV estimates is close to the true value. The corresponding standard errors again seem small, considering that the datasets only have 100 observations. The average OLS estimates are again all more than two standard deviations from their true values, indicating again the necessity of accounting for endogeneity in the regression.

4   Discrete choice and generalized entropy

According to Theorem 2 there exists a generalized entropy that leads to a demand system that is not consistent with any ARUM. This section establishes that the class of demand systems (3) that can be created using generalized entropy includes all demands systems derived from ARUM. The class of generalized entropy demands is thus strictly larger than the class of ARUM demands.

We consider ARUM with utilities $v_j + \varepsilon_j$, $j \in \{1, ..., J\}$, where $v = (v_1, ..., v_J)$ is deterministic and $\varepsilon = (\varepsilon_1, ..., \varepsilon_J)$ is a vector of random utility residuals. The joint distribution of $\varepsilon = (\varepsilon_1, ..., \varepsilon_J)$ is absolutely continuous with finite means and independent of v. Suppose for simplicity that ε is supported on all of $\mathbb{R}^J$. Each consumer draws a realization of ε and chooses the alternative $\chi = \operatorname{argmax}_j \{v_j + \varepsilon_j\}$ with the maximum utility, such that ε_χ is the residual of the maximum utility alternative. The expected maximum utility is denoted

$$G(v) = E(v_\chi + \varepsilon_\chi). \tag{12}$$

We denote the vector of choice probabilities as $P(v) = (P_1(v), ..., P_J(v))$, where $P_j(v) = P(\chi = j)$. It is well known that $P(v) = \nabla G(v)$ (McFadden, 1981). All choice probabilities are everywhere positive since ε has full support. The following lemma collects some properties of G and ε_χ.

Lemma 3 *The function G is convex and finite everywhere. G has the homogeneity property that $G(v + c) = G(v) + c$ for any $c \in \mathbb{R}$, and G is twice continuously differentiable. Furthermore, G is given in terms of the expected residual of the maximum utility alternative by*

$$G(v) = P(v) \cdot v + E(\varepsilon_\chi | v).$$

When the function G is convex and finite, it is also continuous and closed. Define

$$H(e^v) = \nabla_v \left(e^{G(v)} \right). \tag{13}$$

It follows directly from this definition that

$$\nabla G(v) = \frac{H(e^v)}{1 \cdot H(e^v)}. \tag{14}$$

In the case of the multinomial logit model, $G(v) = \ln \sum_{j=1}^{J} e^{v_j}$, $H(e^v) = e^v$, such that (14) is the well known expression for the probabilities of that model.

Lemma 4 is essentially the content of the appendix in Berry (1994). However, the proof in Berry relies on the existence of an outside option. The present proof

does not require an outside option to be present. The proof of Lemma 4 uses Lemma 1, which allows it to be quite short.

Lemma 4 *The function H defined by $H(e^v) = \nabla_v\left(e^{G(v)}\right)$ is invertible.*

The invertibility of H allows us to define

$$S(q) = H^{-1}(q). \tag{15}$$

Let

$$G^*(q) = \sup_v \{q \cdot v - G(v)\} \tag{16}$$

be the convex conjugate of G (Rockafellar, 1970, p. 104). Theorem 4 provides an explicit form for $G^*(q)$, which underlies the findings that we present below. The function $G^*(q)$ is finite only on the unit simplex Δ, the set of probability vectors.

Theorem 4 *The convex conjugate of the expected maximum utility $G(v)$ is*

$$G^*(q) = \begin{cases} q \cdot \ln S(q), & q \in \Delta \\ +\infty, & q \notin \Delta. \end{cases}$$

Moreover, $G(v) = \sup_q \{q \cdot v - G^*(q)\}$ and $E(\varepsilon_\chi | v) = -G^*(q)$ when $q = \nabla G(v)$.

When ε is an i.i.d. extreme value type 1 vector, then $G(v) = \ln(1 \cdot e^v)$, while $-G^*(q) = -q \cdot \ln q$ is the Shannon entropy (Shannon, 1948). This shows that $-G^*(q)$ is a generalization of entropy. We shall explore some properties of this generalization.

The generalization of entropy $-G^*(q)$ is concave, since G^* is the convex conjugate of a convex function. It has maximum where $0 \in \partial G^*(q)$ or equivalently where $\partial G^*(q) = \{v | v = (c, ..., c), c \in \mathbb{R}\}$. Hence it is maximal at the probability vector corresponding to vectors v that are constant across choice alternatives in the ARUM and do not affect the discrete choice. This is consistent with the interpretation of entropy as a measure of the expected surprise associated with a distribution.

The Shannon entropy is always positive. The generalization of entropy $-G^*(q)$

may take any value, but it is necessarily positive when the random components have zero mean - this is a direct consequence of Jensen's inequality.

Proposition 6 *If $E(\varepsilon_j) = 0$ for all j in an ARUM, then the corresponding generalized entropy is always non-negative: $-G^*(q) \geq 0, q \in \Delta$.*

We now turn to establishing the relation between ARUM and generalized entropy. The following two lemmas are used to show that a function S derived from an ARUM is a flexible generator as defined in Section 2.

Lemma 5 *The function $S = H^{-1}$ is continuous, homogenous of degree 1, and satisfies Condition 3.*

We note by Lemmas 4 and 5 that an S derived from an ARUM via (15) is a flexible generator. The ARUM demand (14) is the same as the demand (3) resulting from maximization of utility (1). Then, by Theorem 4, we have proved

Theorem 5 *Let G^* be the convex conjugate of an ARUM surplus function $G(v) = E\max_j\{v_j + \varepsilon_j\}$. Then $-G^*$ is a generalized entropy. The ARUM demand equals the utility maximizing demand in Theorem 1.*

Section 2.3 provided an example of a generalized entropy that is not the convex conjugate of an ARUM surplus function.

4.1 Application to discrete choice data

We shall consider how to apply the generalized entropy model to microdata with observations of discrete choices. Such data are commonly available and provide the opportunity for incorporating individual specific information. The associated cost is that it is not possible to estimate microdata models merely by regression in the same way as with market level data. This section establishes an algorithm for computing the likelihood and applies this in an example to estimate a model using maximum likelihood. The ability to compute the likelihood is also useful with bayesian methods in combination with the Metropolis-Hastings algorithm.

We take as a starting point that individuals choose good j with probability q_j satisfying $v = \ln S(q) + c$ for some flexible generator S and with $c \in \mathbb{R}$ ensuring

that probabilities sum to 1. If the generalized entropy in utility (1) is the convex conjugate of an ARUM surplus function, then q are simply the corresponding discrete choice probabilities. Generalized entropies that are not ARUM consistent may still correspond to nonadditive random utility models, i.e. models where utilities are not just sums but more general functions of v_j and ε_j (Matzkin, 2007). Alternatively, individuals could be seen as making random choices with probabilities that are the result of utility maximization (Fudenberg et al., 2014).

We will consider estimation by maximum likelihood. This requires us to compute the likelihood q given v and we hence need a way to invert S that is feasible within a maximum likelihood routine. The following theorem indicates how the likelihood may computed by using an iterative process to solve a fixed point problem. We use the Kullback and Leibler (1951) distance function $d_r(q) = r \cdot \ln\left(\frac{r}{q}\right)$ to evaluate the distance from the fixed point r to some q. This is a convex function with minimum at r with $d_r(r) = 0$. Hence $d_r(q)$ will be larger the further q is from r.

Theorem 6 *Let S be the flexible generator defined in Proposition 2 and let $r \in \Delta$ satisfy $v = \ln S(r) + c$ for some $c \in \mathbb{R}$. Then the mapping*

$$w(q) = \left\{ \frac{q_i e^{v_i} / S^{(i)}(q)}{\sum\limits_j q_j e^{v_j} / S^{(j)}(q)} \right\} \tag{17}$$

has r as unique fixed point and iteration of (17) from any starting point in Δ converges to r.

If S has the form

$$S^{(j)}(q) = q_j^{\mu_0} \prod_{\{g \in \mathcal{G} | j \in g, g \neq \{j\}\}} q_g^{\mu_g} \tag{18}$$

for some $\mu_0 > 0$, then $d_r(w(q)) \leq (1 - \mu_0) d_r(q)$.

Theorem 6 then shows that iteration of (17) will always converge to the fixed point. Intuitively, the numerator of (17) adjusts each q_i in the direction that makes $v = \ln S(q) + c$ true, while the denominator ensures that $1 \cdot w(q) = 1$. The second

Table 3: Maximum likelihood estimates in discrete choice simulation with cross-nested model

	α	β	μ_1	μ_2
True parameters	0.500	0.500	0.200	0.500
Avg. estimates	0.498	0.498	0.208	0.495
Std.dev.	0.050	0.050	0.043	0.055

part of the theorem concerns the special case when the flexible generator is an average of the identity with something else. Beginning from q^0 and iterating such that $q^n = w(q^{n-1})$, $n \geq 1$ the theorem shows that $d_r(q^n) \leq (1-\mu_0)^n d_r(q^0)$, which means that the distance to the fixed point decreases exponentially

A question is now how well it is possible to recover parameters underlying utility from the observation of discrete choices. We have investigated this in a simulation experiment where we have simulated data from the cross-nested structure of Section 3.1. We do not include the outside option as we have a situation in mind where we observe the choices of consumers who buy one of the varieties of some good under consideration. Utilities are specified as $v_j = \alpha x_{1j} + \beta x_{1j}x_2$, where x_{1j} represents an alternative specific characteristic, while x_2 represents individual specific variation. We performed 100 replications with 1000 individuals in each, each individual selects 1 among the 9 alternatives in the model with probabilities q, where $\ln S(q) = v + c$. The independent variables were generated as i.i.d. standard normal. The likelihood was computed using Theorem 6 and was maximized numerically.[8] The results are summarized in Table 3. As in the previous simulations in this paper, we find that the true parameters are well recovered.

5 Concluding remarks

This paper has introduced the concepts of generalized entropy and flexible generators and used them to derive a general family of demand systems. General rules for constructing demand systems have been provided along with some specific examples and it has been shown how these models may be estimated using either market share or individual level data.

---

[8]Using BFGS with numerical derivatives.

We believe that generalized entropy models may be useful in a range of circumstances. One example that we have mentioned is the demand for automobiles (e.g. Berry et al., 1995; Goldberg and Verboven, 2001; Train and Winston, 2007). The number of varieties of new cars is large and there are likely complex substitution patterns that may be accounted for using flexible generators. Another application area characterized by a large number of alternative "products" is spatial models, where flexible generators may be used to describe spatial correlations, for example in models of equilibrium sorting (Kuminoff et al., 2013). Yet another area where generalized entropy models may be useful are matching models (Salanié and Galichon, 2015), where the range of possible models could be extended. It would also be of interest to develop generalized entropy models in the context of dynamic discrete choice models (Chiong et al., 2015). We hope that the family of demand systems provided here will stimulate future empirical work.

The generalized entropy model extends to the case where the vector v is random with each consumer having some realization of v. Then demand conditional on v still has the form (3) and the expected demand is the expectation of (3). This is analogous to the mixed logit model (McFadden and Train, 2000). Moreover, both in the present case and in the mixed logit, the presence of the expectation implies that the explicit inversion in Theorem 1 does not carry through when v is random.

The nesting device we use to create flexible generators does not exhaust all possibilities. One possibility that we have not explored, for example, is to combine our nesting device with the idea that membership of a nest may be partial. There is thus scope for finding more flexible generators with properties that may be useful in specific circumstances.

Acknowledgements

References

Anderson, S. P., de Palma, A. and Thisse, J.-F. (1988) A Representative Consumer Theory of the Logit Model *International Economic Review* **29**(3), 461.

Berry, S. (1994) Estimating discrete-choice models of market equilibrium *The RAND Journal of Economics* **25**(2), 242–262.

Berry, S. and Haile, P. A. (2014) Identification in Differentiated Products Markets Using Market Level Data *Econometrica* **82**(5), 1749–1797.

Berry, S., Levinsohn, J. and Pakes, A. (1995) Automobile Prices in Market Equilibrium *Econometrica* **63**(4), 841–890.

Birkhoff, G. (1946) Tres observaciones sobre el algebra lineal *Universidad Nacional de Tucuman, Revista. Serie A* **5**, 147–151.

Chiong, K. X., Galichon, A. and Shum, M. (2015) Duality in dynamic discrete choice models *Available at SSRN: http://ssrn.com/abstract=2700773 or http://dx.doi.org/10.2139/ssrn.2700773* .

Fosgerau, M., McFadden, D. and Bierlaire, M. (2013) Choice probability generating functions *Journal of Choice Modelling* **8**, 1–18.

Fosgerau, M. and McFadden, D. L. (2012) A theory of the perturbed consumer with general budgets *Working Paper 17953* National Bureau of Economic Research.

Fudenberg, D., Iijima, R. and Strzalecki, T. (2014) Stochastic Choice and Revealed Perturbed Utility *Working Paper* .

Goldberg, P. K. and Verboven, F. (2001) The Evolution of Price Dispersion in the European Car Market *The Review of Economic Studies* **68**(4), 811–848.

Knittel, C. R. and Metaxoglou, K. (2014) Estimation of Random-Coefficient Demand Models: Two Empiricists' Perspective *Review of Economics and Statistics* **96**(1), 34–59.

Kullback, S. and Leibler, R. A. (1951) On Information and Sufficiency *The Annals of Mathematical Statistics* **22**(1), 79–86.

Kuminoff, N. V., Smith, V. K. and Timmins, C. (2013) The New Economics of Equilibrium Sorting and Policy Evaluation Using Housing Markets *Journal of Economic Literature* **51**(4), 1007–62.

Mattsson, L.-G. and Weibull, J. W. (2002) Probabilistic choice and procedurally bounded rationality *Games and Economic Behavior* **41**(1), 61–78.

Matzkin, R. L. (2007) Chapter 73 Nonparametric identification *in* J. J. H. a. E. E. Leamer (ed.), *Handbook of Econometrics* Vol. 6, Part B Elsevier pp. 5307–5368.

McFadden, D. (1978) Modelling the choice of residential location *in* A. Karlquist, F. Snickars and J. W. Weibull (eds), *Spatial Interaction Theory and Planning Models* North Holland Amsterdam pp. 75 –96.

McFadden, D. (1981) Econometric Models of Probabilistic Choice *in* C. Manski and D. McFadden (eds), *Structural Analysis of Discrete Data with Econometric Applications* MIT Press Cambridge, MA, USA pp. 198–272.

McFadden, D. and Train, K. (2000) Mixed MNL Models for discrete response *Journal of Applied Econometrics* **15**, 447–470.

Melo, E. (2012) A representative consumer theorem for discrete choice models in networked markets *Economics Letters* **117**(3), 862–865.

Mirsky, L. (1958) Proofs of two theorems on doubly stochastic matrices *Proceedings of the American Mathematical Society* **9**, 371–374.

Rockafellar, R. (1970) *Convex analysis* Princeton University Press Princeton, N.J.

Ruzhansky, M. and Sugimoto, M. (2014) On global inversion of homogeneous maps *Bulletin of Mathematical Sciences* pp. 1–6.

Salanié, B. and Galichon, A. (2015) Cupid's Invisible Hand: Social Surplus and Identification in Matching Models *Working Paper, Columbia University* .

Shannon, C. E. (1948) A Mathematical Theory of Communication *Bell System Technical Journal* **27**(3), 379–423.

Train, K. E. and Winston, C. (2007) Vehicle choice behavior and the declining market share of u.s. automakers *International Economic Review* **48**(4), 1469–1496.

Verboven, F. (1996) The nested logit model and representative consumer theory *Economics Letters* **50**(1), 57–63.

Appendixes

A   Proofs for Section 2

Proof of Theorem 1.   Form the Lagrangian

$$\Lambda\left(q,\lambda\right)=\tau y+q\cdot v-q\cdot\ln S\left(q\right)+\lambda\left(1-1\cdot q\right).$$

The first-order conditions for $(q_1,...,q_J)$ are

$$0=\frac{\partial\Lambda}{\partial q_k}=v_k-\ln S^{(k)}\left(q\right)-\sum_{j=1}^{J}q_j\frac{d\ln S^{(j)}\left(q\right)}{dq_k}-\lambda,$$

resulting by Condition 3 in

$$S\left(q\right)=e^{v-\kappa-\lambda}>0.$$

The homogeneity of S implies homogeneity of $H=S^{-1}$ and then

$$q=H\left(e^{v-\kappa-\lambda}\right)=e^{-(\kappa+\lambda)}H\left(e^{v}\right).$$

The constraint $1\cdot q=1$ implies that $e^{\kappa+\lambda}=1\cdot H\left(e^{v}\right)$ such that any solution to the first-order conditions satisfies

$$q=\frac{H\left(e^{v}\right)}{1\cdot H\left(e^{v}\right)}\tag{19}$$

and thus q is uniquely determined.

Existence of a solution is established as follows: Existence can fail only if the denominator in (19) is zero; but since the $H^{(j)}\left(e^{v}\right)$ are non-negative, this can only occur if $H^{(j)}\left(e^{v}\right)=0$ for all j; this implies in turn by invertibility and homogeneity of S that $e^{v}=0$, which is a contradiction. By Condition (2), the utility $u\left(q\right)$ is concave, and hence the solution (19) to the first-order conditions is a global maximum.

To prove the second part of the theorem, note first that if q is an interior so-

lution to the utility maximization problem then it satisfies equation (3), which implies that

$$\ln S(q) + \ln(1 \cdot H(e^v)) = v.$$

Conversely, if $v = \ln S(q) + c$, then q solves (3). ∎

Proof of Lemma 1. This follows from Theorem 2.4 in Ruzhansky and Sugimoto (2014) upon noting that S may be extended to

$$f(x) = \begin{cases} S(x), & x \in (0, \infty)^J \\ x, & x \in \mathbb{R}^J \setminus (0, \infty)^J. \end{cases}$$

Note that $A \equiv \mathbb{R}^J \setminus (0, \infty)^J$ is closed , f is C^1 on $\mathbb{R}^J \setminus A$ with $\det J_f \neq 0$ on $\mathbb{R}^J \setminus A$, f is continuous and injective on A, and $\mathbb{R}^J \setminus f(A)$ is simply connected. It is also the case that $f(\mathbb{R}^J \setminus A) \subset \mathbb{R}^J \setminus f(A)$. Let $\{x_n\} \subseteq (0, \infty)^J$ with $\|x_n\| \to \infty$. Then $\|S(x_n)\| = \|x_n\| \left\| S\left(\frac{x_n}{\|x_n\|}\right) \right\| \geq \|x_n\| \inf_{q \in \Delta} S(q) \to \infty$. Since f satisfies the conditions in the Ruzhansky and Sugimoto (2014) theorem, S is invertible. ∎

Proof of Lemma 2. (Averaging) Conditions 1-3 are easily verified. We shall verify that T is invertible using Lemma 1. Since $T_k^{(j)}(q) \geq q_j$, also $S^{(j)}(q) \geq q_j$, and then $\inf_q \|S(q)\| \geq J^{-1} > 0$, which is the first requirement in Lemma 1.

The Jacobian of $\ln S$ is

$$J_{\ln S} = \sum_{k=1}^{K} \alpha_k J_{\ln T_k}.$$

Then $J_{\ln S}$ is positive definite and hence its determinant is positive. The Jacobian $J_S = diag\left\{ S^{(1)}(q)^{-1}, ..., S^{(J)}(q)^{-1} \right\} \cdot J_{\ln S}$ also has positive determinant, which is the second requirement in Lemma 1. ∎

Proof of Proposition 2. (General nesting) Condition 1 follows directly. Condition 2 follows by noting that $\Omega(q)$ is a linear combination of functions of the type

$-t \ln t$ and that $t \to -t \ln t$ is strictly concave when $t > 0$. Finally,

$$
\begin{aligned}
\sum_{j=1}^{J} q_j \frac{d \ln S^{(j)}(q)}{dq_k} &= \sum_{j=1}^{J} q_j \frac{\sum_g \mu_g 1_{\{j \in g\}} 1_{\{k \in g\}} \partial \ln (q_g)}{\partial q_k} \\
&= \sum_{g \in \mathcal{G}} \mu_g 1_{\{k \in g\}} \sum_{j=1}^{J} \frac{q_j 1_{\{j \in g\}}}{q_g} = 1
\end{aligned}
$$

showing that Condition 3 holds as required.

We have

$$
S^{(j)}(q) = \prod_{\{g \in \mathcal{G} | j \in g\}} q_g^{\mu_g} \geq \prod_{\{g \in \mathcal{G} | j \in g\}} q_j^{\mu_g} = q_j.
$$

The Jacobian of $\ln S$ has elements jk

$$
\sum_{\{g \in \mathcal{G} | j \in g, k \in g\}} \mu_g \frac{1}{q_g},
$$

such that it is symmetric and positive semidefinite. If it is positive definite, then by Lemma 2 S has an inverse and is a flexible generator. ∎

Proof of Proposition 3. (Invertible nesting) Observe that (6) may be written in matrix form as $\ln S(q) = MW \ln (M^\top q)$. Then

$$
\begin{aligned}
\ln S(q) &= v \Leftrightarrow \\
q &= \left(M^\top\right)^{-1} \exp \left(W^{-1} M^{-1} v\right).
\end{aligned}
$$

Hence S has an inverse and it follows from Proposition 2 that S is a flexible generator. ∎

Proof of Proposition 4. (Transformation) We shall verify Conditions 1-4. Observe that S defined by (8) is continuous and that for $\alpha > 0$,

$$
S(\alpha q) = \exp \left(\ln \alpha + A^\top \left[\ln (T(Aq))\right] + m\right) = \alpha S(q),
$$

since columns of A sum to 1. This verifies Condition 1.

Let $\Omega_T(q) = -q \cdot \ln T(q)$; this is concave on Δ by assumption. Note that for

$q \in \Delta$,

$$
\begin{aligned}
-\Omega_T (Aq) + q \cdot m &= Aq \cdot \ln T (Aq) + q \cdot m \\
&= \sum_i \left(\sum_j a_{ij} q_j \right) \ln T^{(i)} (Aq) + q \cdot m \\
&= \sum_j q_j \left[\sum_i a_{ij} \ln T^{(i)} (Aq) \right] + q \cdot m \\
&= q \cdot \left(A^\top [\ln (T (Aq))] + m \right) \\
&= q \cdot \ln S (q) .
\end{aligned}
$$

Hence, $-\Omega_T (Aq) + q \cdot m = q \cdot \ln S (q)$. The transformation $q \to \Omega_T (Aq) - q \cdot m$ is concave on Δ (Rockafellar, 1970, Thm. 12.3) and then so is $-q \cdot \ln S (q)$.

The next step is to verify Condition 3. The condition holds by assumption for Ω_T, and may be expressed as $-\nabla \Omega_T (q) = \ln T (q) + 1$. Now, with $\Omega_S (q) = -q \cdot \left(A^\top [\ln (T (Aq))] + m \right) = -(Aq)^\top \cdot [\ln (T (Aq))] - q \cdot m$, we see that

$$
\begin{aligned}
-\nabla \Omega_S (q) &= A^\top (-\nabla \Omega_T (Aq)) + m \\
&= A^\top (\ln T (Aq) + 1) + m \\
&= \ln S (q) + 1
\end{aligned}
$$

as required.

Finally, we shall verify the invertibility Condition 4 by solving the equation $\ln T (q) = v$.

$$
\begin{aligned}
\ln T (q) &= A^\top [\ln (S (Aq))] + m = v \Leftrightarrow \\
q &= A^{-1} H \left(\exp \left[\left(A^\top \right)^{-1} (v - m) \right] \right) .
\end{aligned}
$$

Thus, T is invertible. This completes the proof. ∎

Proof of Proposition 5. (Nesting) Consider the function S defined by (9). Condition 1 is clearly satisfied. Generalized entropy corresponding to S is

$$
\Omega (q) = -q^1 \cdot \ln T_1 \left(\frac{q^1}{1 \cdot q^1} \right) - \left(1 \cdot q^1, q^2 \right) \cdot \ln T_2 \left(1 \cdot q^1, q^2 \right), q \in \Delta,
$$

which is concave and then Condition 2 is satisfied.

For $q = (q^1, q^2) = (q_1, ..., q_{J_1+J_2-1})$, and with $r = \frac{q^1}{1 \cdot q^1}$, we have

$$
\sum_{j=1}^{J_1+J_2-1} q_j \frac{\partial \ln S^{(j)}(q)}{\partial q_k}
$$

$$
= 1_{\{k \leq J_1\}} \sum_{j=1}^{J_1} q_j^1 \frac{\partial \ln T_1^{(j)} \left(\frac{q^1}{1 \cdot q^1} \right)}{\partial q_k} + (1 \cdot q^1) \frac{\partial \ln T_2^{(1)}(1 \cdot q^1, q^2)}{\partial q_k}
$$

$$
+ \sum_{j=2}^{J_2-1} q_{J_1+j} \frac{\partial \ln T_2^{(j)}(1 \cdot q^1, q^2)}{\partial q_k}
$$

$$
= 1_{\{k \leq J_1\}} (1 \cdot q^1) \sum_{j=1}^{J_1} r_j \sum_{i=1}^{J_1} \frac{\partial \ln T_1^{(j)}(r)}{\partial r_i} \frac{\partial r_i}{\partial q_k} + 1
$$

$$
= 1_{\{k \leq J_1\}} (1 \cdot q^1) \sum_{j=1}^{J_1} \left[r_j \sum_{i=1}^{J_1} \frac{\partial \ln T_1^{(j)}(r)}{\partial r_i} \left(1_{\{i=k\}} \frac{1}{1 \cdot q^1} - \frac{r_i}{1 \cdot q^1} \right) \right] + 1
$$

$$
= 1_{\{k \leq J_1\}} (1 \cdot q^1) \sum_{i=1}^{J_1} \left[\left(1_{\{i=k\}} \frac{1}{1 \cdot q^1} - \frac{r_i}{1 \cdot q^1} \right) \sum_{j=1}^{J_1} r_j \frac{\partial \ln T_1^{(j)}(r)}{\partial r_i} \right] + 1
$$

$$
= 1_{\{k \leq J_1\}} (1 \cdot q^1) \sum_{i=1}^{J_1} \left[\left(1_{\{i=k\}} \frac{1}{1 \cdot q^1} - \frac{r_i}{1 \cdot q^1} \right) \right] + 1
$$

$$
= 1,
$$

which is Condition 3.

Finally, to show that S is invertible, consider the equation $\ln S(q^1, q^2) = (v^1, v^2)$. Let $q^1 = T_1^{-1}\left(e^{v^1} \right)$, $r = 1 \cdot q^1$, and let $(rs, q^2) = T_2^{-1}\left(r, e^{v^2} \right)$. Then for $j \leq J_1$

$$
\begin{aligned}
\ln S^{(j)}\left(sq^1, q^2 \right) &= \ln T_1^{(j)}\left(\frac{q^1}{r} \right) + \ln T_2^{(1)}\left(rs, q^2 \right) \\
&= \ln T_1^{(j)}\left(q^1 \right) - \ln r + \ln T_2^{(1)}\left(rs, q^2 \right) \\
&= v_j^1 - \ln r + \ln T_2^{(1)}\left(rs, q^2 \right) = v_j^1.
\end{aligned}
$$

For $j > J_1$ we have

$$
\ln S^{(j)}\left(sq^1, q^2 \right) = \ln T_2^{(j-J_1)}\left(rs, q^2 \right) = v_j^2.
$$

Then

$$S^{-1}\left(e^{v^1}, e^{v^2}\right) = \left(sq^1, q^2\right)$$
$$= \left(sT_1^{-1}\left(e^{v^1}\right), q^2\right)$$

and Condition 4 is satisfied. Thus, S is a flexible generator. ∎

B   Proofs for Section 4

Proof of Lemma 3. Fosgerau et al. (2013) establishes convexity and finiteness of G as well as the homogeneity property and the existence of all mixed partial derivatives up to order J. This also implies that all second order mixed partial derivatives are continuous, since $J \geq 3$.

The existence of derivatives G_{ii} is established from the homogeneity property that $G_j(v + c) = G_j(v), j = 1,,,,., J$. Consider G_{11} at no loss of generality and observe that

$$\frac{G_1\left(v_1 + c, v_2, ..., v_J\right) - G_1\left(v_1, v_2, ..., v_J\right)}{c}$$
$$= \frac{G_1\left(v_1, v_2 - c, ..., v_J - c\right) - G_1\left(v_1, v_2, ..., v_J\right)}{c}$$
$$\underset{c \to 0^+}{\to} -\sum_{j \neq 1} G_{1j}(v) = G_{11}(v),$$

which means that G_{11} exists. Furthermore, $G_{1j}, j > 1$ are continuous and hence so is G_{11}.

Let χ be the index of the chosen alternative. The last statement of the lemma follows using the law of iterated expectations since

$$G(v) = \sum_j E\left(\max_j \{v_j + \varepsilon_j\} | \chi = j, v\right) P_j(v)$$
$$= \sum_j \left(v_j + E\left(\varepsilon_\chi | \chi = j, v\right)\right) P_j(v)$$
$$= P(v) \cdot v + E\left(\varepsilon_\chi | v\right).$$

∎

Proof of Lemma 4. We shall make use of Lemma 1 applied to H. The Jacobian of $v \to H(e^v)$ is $\{e^{G(v)}G_i(v)G_j(v)\} + \{e^{G(v)}G_{ij}(v)\}$. The first matrix is positive definite since all choice probabilities are positive, the second matrix is positive semidefinite due to the convexity of G, hence this matrix is everywhere positive definite and then the Jacobian determinant of $v \to H(e^v)$ never vanishes. This implies in turn that the Jacobian determinant of the composition $y \to \ln y \to H(y)$ never vanishes. It remains to show that $\inf_{y \in \Delta} \|H(y)\| > 0$. But $y \in \Delta$ implies that

$$
\begin{aligned}
\|H(y)\| &= e^{G(\ln y)} \|\nabla G(\ln y)\| \\
&\geq e^{E \max_j \{\ln y_j + \varepsilon_j\}} J^{-1/2} \\
&\geq e^{\max_j \{\ln y_j + E\varepsilon_j\}} J^{-1/2} \\
&= \max_j \{y_j e^{E\varepsilon_j}\} J^{-1/2} \\
&\geq \left\| (y_1 e^{E\varepsilon_1}, ..., y_J e^{E\varepsilon_J}) \right\| J^{-1} \\
&\geq \left(\sum_{j=1}^J e^{-2E\varepsilon_j} \right)^{-1} J^{-1} > 0,
\end{aligned}
$$

where we first used that ∇G is on the unit simplex, second that the max operation is convex, third that the sup-norm bounds the euclidean norm, and fourth that the minimum of $\left\| (y_1 e^{E\varepsilon_1}, ..., y_J e^{E\varepsilon_J}) \right\|$ on the unit simplex is attained at $y_j = e^{-2E\varepsilon_j} \left(\sum_{k=1}^J e^{-2E\varepsilon_k} \right)^{-1}, j = 1, ..., J.$ ∎

Proof of Theorem 4. We first evaluate $G^*(q)$. If $1 \cdot q \neq 1$, then

$$
q \cdot (v + \gamma) - G(v + \gamma) = q \cdot v - G(v) + (1 \cdot q - 1)\gamma,
$$

which can be made arbitrarily large by changing γ and hence $G^*(q) = \infty$. Next consider q with some $q_j < 0$. $G(v)$ decreases towards a lower bound denoted $G(-\infty, v_{-j})$ as $v_j \to -\infty$. Then $q \cdot v - G(v)$ increases towards $+\infty$ and hence G^* is $+\infty$ outside the unit simplex Δ.

For $q \in \Delta$, we solve the maximization problem (16) noting that we may fix $1 \cdot v = 0$. Maximize then the Lagrangian $q \cdot v - G(v) - \lambda(1 \cdot v)$ with first-order

conditions $0 = q_j - G_j(v) - \lambda$, which lead to $\lambda = 0$. Then

$$
\begin{aligned}
q &= \nabla_v G(v) \Leftrightarrow \\
q e^{G(v)} &= \nabla_v \left(e^{G(v)}\right) = H(e^v) \Leftrightarrow \\
S(q) e^{G(v)} &= e^v \Leftrightarrow \\
\ln S(q) + G(v) &= v \Rightarrow \\
q \cdot \ln S(q) + G(v) &= q \cdot v.
\end{aligned}
$$

Inserting this into (16) leads to the desired result.

G is convex and closed and hence G is the convex conjugate of G^* (Rockafellar, 1970, Thm. 12.2), this is the next assertion of the theorem. Finally, for $q = \nabla G(v)$, a fundamental result of convex analysis (Rockafellar, 1970, Thm. 23.5) states that $G(v) + G^*(q) = v \cdot q$, which may be combined with (12) to yield the final statement of the theorem. ■

Proof of Proposition 6. Note that the maximum is a convex function, such that Jensen's inequality applies. Then, for $q = \nabla G(v)$,

$$
\begin{aligned}
-G^*(q) &= E \max_j (v_j + \varepsilon_j) - v \cdot q \\
&\geq \max_j E(v_j + \varepsilon_j) - v \cdot q \geq 0.
\end{aligned}
$$

■

Proof of Lemma 5. Continuity of S follows from continuity of the partial derivatives of G, which is immediate from the definition.

Homogeneity of S is equivalent to homogeneity of H. Using the homogeneity property of G,

$$
S^{-1}(\lambda e^v) = \nabla_v \left(e^{G(v + \ln \lambda)}\right) = \lambda \nabla_v \left(e^{G(v)}\right) = \lambda S^{-1}(e^v),
$$

which shows that H and hence S are homogenous of degree 1.

The requirement that $\sum_{j=1}^{J} q_j \frac{d \ln S^{(j)}(q)}{dq_k} = \kappa$ on $\mathrm{relint}(\Delta)$ may be expressed in matrix notation in terms of the Jacobian $J_{\ln S}(q_1, ..., q_J)$ of $\ln S$ as $(q_1, ..., q_J) \cdot J_{\ln S}(q) = \kappa(1, ..., 1)$. We take $\kappa = 1$.

With $v = \ln S(q)$ and noting that $(\ln S)^{-1}(v) = H(e^v)$, the requirement is equivalent to

$$(q_1, ..., q_J) = (q_1, ..., q_J) \cdot J_{\ln S}(q) \cdot J_{(\ln S)^{-1}}(v) = (1, ..., 1) \cdot J_{H(e^v)}(v).$$

Now,

$$
\begin{aligned}
(1, ..., 1) \cdot J_{H(e^v)}(v) &= (1, ..., 1) \cdot \left\{ e^{G(v)} G_j(v) G_k(v) + e^{G(v)} G_{jk}(v) 1_{\{j=k\}} \right\} \\
&= (q_1, ..., q_J)
\end{aligned}
$$

as required. We have used first that

$$(q_1, ..., q_J) = H(e^v),$$

and second that $(1, ..., 1) \cdot \{G_{jk}(v)\} = 0$, where the latter assertion follows since $1 = \sum_{j=1}^{J} G_j(v)$. $\blacksquare$

Proof of Theorem 6. Note that $\sum_g \mu_g r_g = \sum_j \sum_{\{g|j \in g\}} \mu_g r_j = 1$. We will first show existence and uniqueness of a fixed point. Note that for $r \in \Delta$:

$$w(r) = \left\{ \frac{r_i e^{v_i}/S^{(i)}(r)}{\sum_j r_j e^{v_j}/S^{(j)}(r)} \right\} = \left\{ \frac{r_i S^{(i)}(r)/S^{(i)}(r)}{\sum_j r_j S^{(j)}(r)/S^{(j)}(r)} \right\} = r,$$

which shows that r is a fixed point. If $q \in \Delta$ is a fixed point, potentially different from r, then $q_i = q_i \left(e^{v_i}/S^{(i)}(q) \right) e^{-c}$, where $e^c = \sum_j q_j e^{v_j}/S^{(j)}(q)$, and then $v = \ln S(r) + c$. The invertibility of S implies that $q = r$ and then the fixed point is unique.

We then need to show that iterations with (17) from any starting point in Δ converges to the fixed point. Define for convenience

$$\pi_j = \frac{S^{(j)}(r)}{S^{(j)}(q)} = \prod_{\{g|j \in g\}} \left(\frac{r_g}{q_g} \right)^{\mu_g}, \quad w_j(q) = \frac{q_i e^{v_i}/S^{(i)}(q)}{\sum_j q_j e^{v_j}/S^{(j)}(q)}.$$

Using that $v = \ln S\left(r\right) + c$ with $c \in \mathbb{R}$ we may rewrite (17) as

$$w_j\left(q\right) = \frac{q_j e^{v_j}/S^{(j)}\left(q\right)}{\sum\limits_i q_i e^{v_i}/S^{(i)}\left(q\right)} = \frac{q_j \frac{S^{(j)}(r)}{S^{(j)}(q)}}{\sum\limits_i q_i \frac{S^{(i)}(r)}{S^{(i)}(q)}} = \frac{q_j \pi_j}{q \cdot \pi}.$$

We will show that $d_r\left(w\left(q\right)\right) \leq d_r\left(q\right)$, with strict inequality when $q \neq r$. This will mean that $w\left(q\right)$ is closer to r than q. Evaluating $d_r\left(w\left(q\right)\right)$ leads to

$$
\begin{aligned}
d_r\left(w\left(q\right)\right) &= r \cdot \ln\left(\frac{r}{w\left(q\right)}\right) \\
&= r \cdot \ln\left(\frac{r}{\left\{\frac{q_j \pi_j}{q \cdot \pi}\right\}}\right) \\
&= d_r\left(q\right) + \ln\left(q \cdot \pi\right) - r \cdot \ln \pi \\
&= d_r\left(q\right) + \ln\left(\sum_j q_j \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}\right) - \sum_j r_j \ln \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}.
\end{aligned}
$$

We thus need to bound the last two terms. Observe first that

$$
\begin{aligned}
\ln\left(\sum_j q_j \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}\right) &= \ln\left(\sum_j q_j \exp \sum_{\{g|j\in g\}} \mu_g \ln\left(\frac{r_g}{q_g}\right)\right) \\
&\leq \ln\left(\sum_j \sum_{\{g|j\in g\}} q_j \mu_g \frac{r_g}{q_g}\right) \\
&= \ln\left(\sum_g \sum_{\{j|j\in g\}} q_j \mu_g \frac{r_g}{q_g}\right) \\
&= \ln\left(\sum_g \mu_g r_g\right) = 0,
\end{aligned}
$$

with strict inequality unless $r_g = q_g$ for all g. Equality would imply $S\left(r\right) = S\left(q\right)$, and further that $r = q$, so we conclude the inequality is strict unless $r = q$.

We also need to bound

$$
-\sum_{j} r_j \ln \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g} = -\sum_{j} \sum_{\{g|j\in g\}} r_j \mu_g \ln \left(\frac{r_g}{q_g}\right)
$$
$$
= -\sum_{g} \mu_g r_g \ln \left(\frac{r_g}{q_g}\right)
$$
$$
= -\sum_{g} \mu_g r_g \ln \left(\frac{\mu_g r_g}{\mu_g q_g}\right) \leq 0,
$$

where the last inequality follows since the term is a Kullback-Leibler distance. We conclude that $d_r(w(q)) \leq d_r(q)$ and that the inequality is strict unless $r = q$.

Now consider a sequence $\{q^n\}$ constructed by iterating (17). Then $d_r(q^n)$ is weakly decreasing and hence $d_r(q^n) \to \rho$ for some $\rho \geq 0$. If $\rho > 0$, then a convergent subsequence can be extracted from $\{q^n\}$ with limit point $\hat{q}$ satisfying $d_r(\hat{q}) = \rho$ by continuity of w. Now $d_r(w(\hat{q})) < \rho$, while there are points from the sequence $\{q^n\}$ arbitrarily close to $\hat{q}$ with $d_r(q^n) > \rho$. This contradicts continuity of w and we conclude that $\rho = 0$ and hence that $q^n \to r$.

If furthermore S has the form (18), then we can improve the bound on $d_r(w(q))$:

$$
d_r(w(q)) - d_r(q) = \ln \left(\sum_{j} q_j \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}\right) - \sum_{j} r_j \ln \prod_{\{g|j\in g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}
$$
$$
\leq -\sum_{j} r_j \ln \left[\left(\frac{r_j}{q_j}\right)^{\mu_0} \prod_{\{g|j\in g,\{j\}\neq g\}} \left(\frac{r_g}{q_g}\right)^{\mu_g}\right]
$$
$$
= -\mu_0 r \cdot \ln \left(\frac{r}{q}\right) - \sum_{j} r_j \sum_{\{g|j\in g,\{j\}\neq g\}} \mu_g \ln \left(\frac{r_g}{q_g}\right)
$$
$$
\leq -\mu_0 d_r(q),
$$

which is the desired result. ∎

Glossary

p	Price vector
P	Probability vector
q, r	Consumption vector of the differentiated good, probability vector
S	Flexible generator
G	Expected maximum utility
G^*	Convex conjugate of G
H	Inverse of S
a	Vector of intrinsic utilities of the differentiated good
Ω	Generalized entropy
v	$\tau\left(a - p\right)$
α, β	Utility parameters
u	Utility
μ	Nesting parameter
Δ	Unit simplex
$\|\cdot\|$	Euclidian norm
τ	Marginal utility of income
χ	Index of chosen alternative